Twitter Thread by Katya



Katya
@ThatKatyaGirl

I have read through the grooming report from the Home Office.

Here, in this thread, I will take you through the mathematical hoops they have jumped through in this deeply flawed analysis.

Thread ■

It is, quite literally, a whitewash. It uses all the standard statistical tricks to avoid concluding a stark and glaring conclusion, which is otherwise plain to see. That there was a sustained, targeted and focussed attack by Pakistani men on white working class and Sikh girls.

Choose limited data points

One way to force a conclusion is to only range over a small and carefully chosen set of data points. In this case, the report used a limited number of use cases which largely ignored the high profile grooming gangs of Pakistani origin.

Only Rotherham was considered in earnest, as it could hardly be ignored as the start of it all.

A typical data trick is to only use one troublesome dataset - generally the most high profile one. This gives plausible denial on accusations of manipulation of source data.

This is what we call “the domain trick”

Present tautologies as statistically relevant

A tautology is a statement which is true by virtue of its logical form. I.e. a statement which is inevitably true. Such statements are not worth declaring as they have no statistical significance.

“Most perpetrators were white” is such a statement.

In a broadly white country, this statement is tautological. The real question is, were other groups present in a statistical

significant way?

I.e. were they over-represented in the the group of perpetrators, given their preponderance in the population as a whole?

The whole point of data analysis is to uncover statistically significance results - not to restate tautologies.

There was no analysis in this report to even attempt this.

This trick does, however, produce lovely quotable headlines which will reassure people of their own anecdotal conclusions.

This is what we call “the pericope trick”

Blame data quality

Whenever you don’t want to use a data item, you can just blame data quality. In this case, they often cited data quality on ethnicity classifications.

The UK ethnicity enumeration is flawed in many areas.

Using it, for example, it is not possible to make a distinction between a native Briton and an East European - or a West African and a South African. These flaws make it a poor tool for a lot of crime analysis.

However, ironically, the enumeration is very strong in the context of South Asian countries. Here ethnicity is broken down into fine detail. Its super easy to gather numbers on Pakistani perpetrators, but the HO claimed data quality on the whole set and didn’t go there.

This is what we call “the quality trick”

Present data ranges as data points

In any given set of statements, there will be a range of values. Each item in the range will then have a corresponding frequency count. So some items might be present once and others hundreds of times.

A common trick though is to simply publish the range as a flat list without frequencies. “Perpetrators come from a range of countries” is this trick in action. It presents the range of nationalities without being concerned with how each presents itself statistically.

The Home Office report lists Portuguese along side Pakistani in a flat list - which, of course, means nothing at all in terms of statistically relevance.

This is what we call “the range trick”

Mixup up data definitions

A final trick I will mention is to mix-up the definition of data types and classes. You can, for example, deliberately use nationality in place of ethnicity when it suits your cause.

Tripping between the two definitions, so that the results always play in favour of your pre-agreed hypothesis.

This is what we call “the predicate trick”

Data will, of course, tell any story you want. When I am tasked with analysing data, I often ask “What do you want it to say?”. Letting data speak for itself, teasing out its hidden truths, allowing it to reveal its story, hardly ever happens - especially in the public sector.

Let me leave the last word to, of all people, Elvis:

“Truth is like the sun. You can shut it out for a time, but it ain't goin' away.”